CLAIMS

1. In a trainable data classifier, a method of forming a measure of difference between first and second data vectors, the method comprising the steps of:

determining an association coefficient of the first and second data vectors; and

forming said measure of difference using said association coefficient.

2. A method according to claim 1 wherein the association coefficient comprises a Jaccard's coefficient.

3. A method according to claim 1 wherein the association coefficient comprises a paired absence measure.

4. A method according to claim 1 further comprising a step of determining a geometric difference between the first and second data vectors, and wherein the step of forming comprises a step of combining said association coefficient and said geometric difference to thereby form said measure of difference.

5. A method according to claim 4 wherein the geometric difference comprises a Euclidean distance.

6. A method according to claim 4 wherein the geometric difference comprises a geometric angle.

7. A method according to claim 4 wherein the step of combining comprises the step of combining the geometric difference and association coefficient in exponential relationship with each other.

8. A method according to claim 7 wherein the step of combining comprises a step of multiplying a function of the geometric difference by an exponent of a function of the association coefficient.

9. A method according to claim 7 wherein the step of combining comprises a step of multiplying a function of the association coefficient by an exponent of a function of the geometric difference.

10. A method according to claim 1 wherein said trainable data classifier comprises a neural network.

11. A method according to claim 1 wherein said first and second data vectors comprise telecommunications account fraud data.

12. A method of retraining a trainable data classifier that has been trained using a plurality of data vectors including a first data vector, the method comprising the steps of:

providing a second data vector;

determining an association coefficient of the first and second data vectors;

forming a measure of conflict between said first and second data vectors using said association coefficient; and

using the second data vector to retrain the data classifier responsive to the measure of conflict.

13. A method according to claim 11 wherein the step of using the second data vector to retrain the data classifier is responsive to a predetermined conflict threshold value.

14. A method according to claim 12 further comprising a step of determining a geometric difference between

the first and second data vectors, and wherein the step of forming comprises a step of combining said association coefficient and said geometric difference to thereby form said measure of conflict.

15. A method of operating a trainable data classifier, said trainable data classifier having been trained using a plurality of training data vectors, said plurality of training data vectors being associated with a plurality of reasons, the method comprising the steps of:

providing an input data vector;

generating an output responsive to the input data vector;

selecting one or more of said training data vectors;

for each selected training data vector:

determining an association coefficient of said input data vector and said selected training data vector, and

forming a measure of difference between said input data vector and said selected training data vector from said association coefficient; and

using said measures of difference to associate at least one of said reasons with said output responsive to said measures of difference.

16. A method according to claim 13 further comprising the step of presenting to a user information indicative of said output, of said at least one of said reasons, and of their association.

17. A method according to claim 13 further comprising the step of using said measures of difference to associate with at least one reason a degree of confidence with which said reason is associated with said input data vector.

18. A method according to claim 15 further comprising a step of determining a geometric difference between said input data vector and said selected training data vector, and wherein the step of forming comprises a step of combining said association coefficient and said geometric difference to thereby form said measure of difference.

19. A method of training a trainable data classifier comprising the steps of:

providing a training data set comprising at least first and second data vectors;

determining an association coefficient of said first and second data vectors;

forming a measure of redundancy between said first and second data vectors from said association coefficient;

modifying said training data set responsive to said measure of redundancy; and

training said trainable data classifier using said modified training data set.

20. A method according to claim 19 wherein the step of forming a measure of redundancy is carried out with reference to a predetermined redundancy threshold value.

21. A method according to claim 19 further comprising the step of discarding one of said first and second data vectors responsive to said measure of redundancy.

22. A method according to claim 19 further comprising a step of determining a geometric difference between said first and second data vectors, and wherein said step of forming comprises a step of combining said association coefficient and said geometric difference to thereby form said measure of redundancy.

23. A data classifier system comprising:

a data classifier operable to provide an output responsive to either of first or second data vectors; and

a data processing subsystem operable to determine an association coefficient of said first and second data vectors, to thereby form a measure of difference between said vectors.

24. A data classifier system according to claim 23 wherein the association coefficient comprises a Jaccard's coefficient.

25. A data classifier system according to claim 23 wherein the association coefficient comprises a paired absences coefficient.

26. A data classifier system according to claim 23 wherein the data processing subsystem is further operable to determine a geometric difference between said first and second data vectors, and to form said measure of difference by combining said association coefficient and said geometric difference.

27. A data classifier system according to claim 26 wherein the geometric difference comprises a Euclidean distance.

28. A data classifier system according to claim 26 wherein the geometric difference comprises a geometric angle.

29. A data classifier system according to claim 26 wherein the data processing subsystem is operable to form said measure of difference by combining said association coefficient and said geometric difference in exponential relationship with each other.

30. A data classifier system according to claim 29 wherein said data processing subsystem is operable to form said measure of difference by multiplying a function of the geometric difference by an exponent of a function of the association coefficient.

31. A data classifier system according to claim 29 wherein said data processing subsystem is operable to form said measure of difference by multiplying a function of the association coefficient by an exponent of a function of the geometric difference.

32. A data classifier system according to claim 23 wherein said data classifier comprises a neural network.

33. An anomaly detection system comprising a data classifier system according to claim 23.

34. An account fraud detection system comprising a data classifier system according to claim 23.

35. A telecommunications account fraud detection system comprising a data classifier system according to claim 23.

36. A network intrusion detection system comprising a data classifier system according to claim 23.

37. Computer software in a machine readable medium for providing at least a part of a data classifier system when executed on a computer system, the software operable to perform the steps of:

receiving first and second data vectors;

determining an association coefficient of the first and second data vectors; and

forming a measure of difference between said

first and second data vectors using said association coefficient.

38. Computer software in a machine readable medium according to claim 37, further operable to perform the step of determining a geometric difference between said first and second data vectors, and to perform the step of forming by carrying out a step of combining said association coefficient and said geometric difference to thereby form said measure of difference.